Supplemental Information

Attentional Enhancement via Selection

and Pooling of Early Sensory Responses

in Human Visual Cortex

Franco Pestilli, Marisa Carrasco, David J. Heeger, and Justin L. Gardner

Supplemental Information

Attentional enhancement via selection and pooling of early sensory responses in human visual cortex

Franco Pestilli, Marisa Carrasco, David J. Heeger and Justin L. Gardner
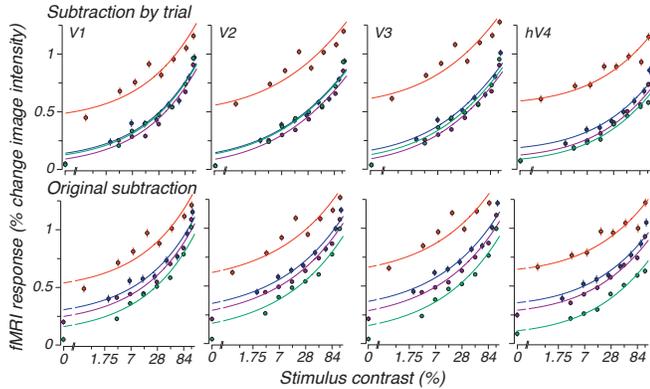
Table of Contents

Supplemental Data
Figure S1. Additional contrast-response functions (see Figure 4)

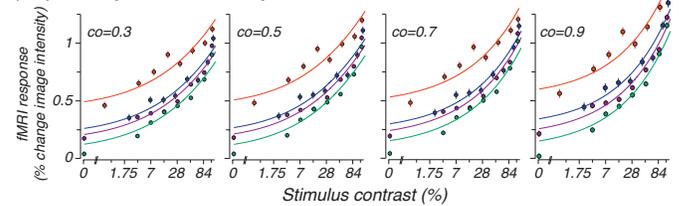**A Contrast response with different 0-response subtraction**
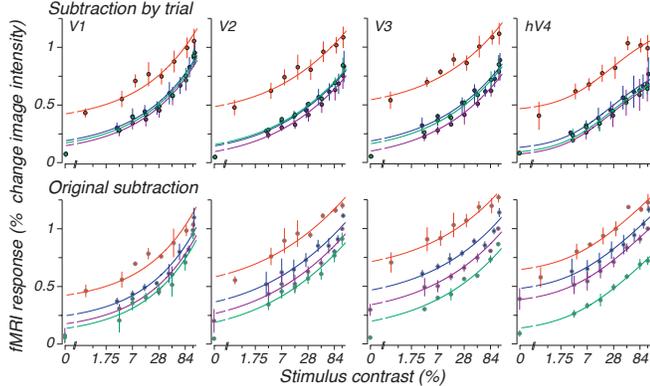
(A.1) V1 responses for one representative observer

B Contrast response with different ROIs

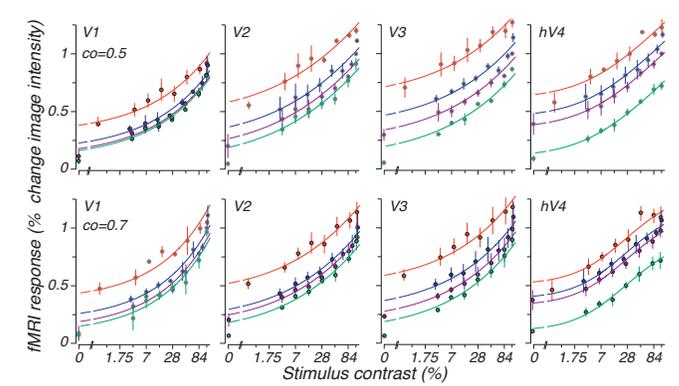**(B.1) V1 responses for one representative observer**

(A.2) Average contrast response

(B.2) Average contrast response

Figure S1

A. Comparing the effects of different zero-response subtractions on the contrast-response function.   A.1 Contrast-response functions for all visual areas (V1, V2, V3, and hV4) for one representative observer. Top row, the response to the zero-contrast stimulus (blank) *focal* non-target was subtracted from all *focal*-target and non-target responses. The response to the zero-contrast stimulus *distributed* non-target was subtracted from all *distributed*-target and non-target responses. Bottom row, the response to the zero-contrast stimulus *focal* non-target was subtracted from all cue conditions.   **A.2** Contrast-response function for all visual areas (V1, V2, V3, and hV4) averaged across observers. Same conventions as **A.1**. See also **Supplemental Experimental Procedures: Minimizing the potential effects of anticipatory BOLD components**.

B. Effect of attention on the contrast-response function at several coherence cut-offs.   B.1 V1 contrast-response function from one representative observer. Each panel displays contrast-response functions averaged across all voxels which met or exceeded the specified coherence cutoff (coherence, co = 0.3, 0.5, 0.7, 0.9) in the phase mapped localizer.   **B.2** Effect of attention on the contrast-response functions (average across observers) at two different coherence cut-off levels (co = 0.5 and 0.7, at which it was possible to reliably define ROIs in each observer and visual area). See also **Supplemental Experimental Procedures: Separating BOLD cortical responses to spatially distant stimuli**.

Figure S2. Additional model fits (See Figure 8)

A Effects of different zero-response subtractions

A.1 By trial

A.2

B Robustness of the models results to the choice of coherence cut-off

B.1

$co=0.3$ $co=0.5$ $co=0.7$

B.2

C Models results with different functional forms of the contrast-discrimination functions

C.1

Equations 3+4   Skewed-Gaussian   Polynomial (order 3)

C.2

Figure S2

A. Robustness of the sensitivity- and selection-model results to removing global anticipatory activity separately for focal and distributed cue trials. Data in these panels are obtained after the response to the focal cue, non-target 0-contrast stimulus was subtracted from the focal cue target stimulus and the response to the distributed cue, non-target 0-contrast stimulus was subtracted from the distributed cue target stimulus. **A.1** Sensitivity model results**.** Sensory noise standard deviations (σ_f) for the focal-cue condition plotted against that of the distributed-cue (σ_f) condition. Shades of color indicate different visual areas V1 to hV4. Symbols indicate different observers, averages across subjects are represented by the larger circles. **A.2** Ratio of noise standard deviation (σ) for focal- to the distributed-cue trials estimated by the selection (left) and the sensitivity (right) models. Same conventions as Fig 8. See also **Supplemental Experimental Procedures: Minimizing the potential effects of anticipatory BOLD components.**

B. Robustness of the sensitivity- and selection-model results to the choice of coherence cut-off.
Panels left to right show results from three coherence cut-off used in the localizer scan to accept voxels for the analysis (0.3, 0.5 and 0.7). Same conventions as above. See also **Supplemental Experimental Procedures: Separating BOLD cortical responses to spatially distant stimuli**.

C. Robustness of the sensitivity- and selection-model results to the choice of functional form used to interpolate data. To ensure that our conclusions about the sensitivity and selection models were not biased by the particular functional form used to interpolate the data, we either used a fit based on Eqs. 3 and 4 (left column) or two simpler and not theoretically motivated equations that fit the data reasonably well and could be used to interpolate the contrast-discrimination performance for fitting the sensitivity model; a skewed gaussian function, (middle column, see Eq. S2) or a third order polynomial, (right column). Each functional form was first fit to the contrast-discrimination data by least-squared error minimization. Then the sensitivity model was used to fit the forms to the fMRI contrast-response functions. These fits were then used to interpolate the contrast-response functions for the selection model. Individual panels have same conventions as above.

Supplemental Experimental Procedures

Behavioral protocol

Observers performed a contrast-discrimination task while fMRI responses were measured. For the duration of each scan (6.5 minutes, 8-10 scans per session), observers maintained fixation on a black-cross (0.5° x 0.5° of visual angle, 1.24 cd/m^2) presented at the center of a gray screen (550 cd/m^2). Each trial lasted 4 s and the intertrial interval was pseudo-randomized (ITI: 4, 6.4 or 8 s). Each trial started with a 1 s cue-interval; an interval long enough to ensure an asymptotic sustained spatial attention effect. During this cue interval (and lasting through the end of the presentation of the discrimination stimuli), either one (focal-cue) or four (distributed-cue) black arrows appeared at fixation (1.24 cd/m^2, 0.7° long and 0.5° wide, appearing 0.5° offset from fixation along the 45° diagonals, see Fig. 2). Single arrow focal-cues indicated with 100% validity the target location. The distributed cues indicated that the target was equally likely to appear at any of the 4 locations. Immediately after, the stimuli appeared at 6° eccentricity in two 600 ms intervals (1st and 2nd intervals) separated by a 200 ms inter-stimulus-interval (ISI). Stimuli were 2-cpd vertical sinusoidal gratings (5° diameter), counter-phase flickering at 5 Hz. A response cue appeared 400 ms after stimulus offset and indicated the target location. Observers were given 1200 ms to respond.

Observers performed a 2-interval-forced-choice discrimination task on these stimuli. On each trial the contrasts of the four stimuli were randomly selected out of eight possible contrasts (pedestal contrasts, 0, 1.75, 3.5, 7, 14, 28, 57 and 84%). One of the four stimuli (target) was presented with a slightly higher contrast (i.e., a small contrast increment, Δc, was added to the pedestal contrast) in one of the two intervals. The rest of the stimuli (non-targets) maintained the same contrast in both stimulus intervals. The response cue indicated the target location. Observers were instructed to report the interval in which the target contrast was higher by pressing the "1" or "2" key on an MRI compatible keypad to indicate that the contrast was higher in the 1st or the 2nd interval. The Δc presented in each trial was set by sixteen independent and randomly interleaved adaptive staircase procedures (QUEST (Watson and Pelli, 1983)), one for each pedestal contrast and cue condition. These independent staircases were set to maintain

observers performance at 76% correct, which under the assumption of normally-distributed and statistically-independent noise without bias, is equivalent to a discriminability (d') of 1. This procedure ensured that we tested many Δc's near threshold contrast and balanced task difficulty across pedestals and cue conditions, which prevented a potential confound between attentional effects and task difficulty.

To minimize spatial and temporal uncertainty, the task design included the following two features. First, to reduce the spatial uncertainty associated with the stimulus locations, we presented four circular frames (6° eccentricity, 5° internal diameter, 5.1° external diameter, 1.24 cd/m^2) around the stimulus locations throughout the duration of the scan, and the response-cue (green arrow, in interval 7) always indicated the target location. Second, to reduce temporal uncertainty, the time from cue onset to stimulus presentation was fixed, and the fixation cross turned white (1100 cd/m^2) to indicate the time intervals during which the stimuli were presented (Fig. 2, intervals 3 and 5).

To ensure that no information about the target location could be gained by comparison with the contrasts of neighboring distracters, their pedestal contrasts were randomized. On each trial we chose the distracter contrasts from the set of pedestal contrasts such that one would have the same contrast as the target, another would have a lower contrast and the third would have a higher contrast. On trials in which the target contrast was set to the highest or lowest contrast in the set of pedestals, we set the distracter contrasts to 28, 57 and 84% and 0, 1.75 and 3.5%, respectively.

To test the effect of high contrast distracters on behavioral performance, we used a slightly modified behavioral protocol. We used pedestal contrasts of 0.044, 7, 14, 28 and 84% contrast. In the control (low contrast distracter) condition, we paired these with distracters chosen to be ¼, ½ and 1.5 times the pedestal contrast (except in the case of the 84% contrast pedestal in which we used an 84% contrast distracter because 1.5 times the pedestal contrast was not achievable). In the high contrast distracter condition the lowest contrast distracter from the control condition was replaced with a distracter of 84% contrast. To prevent observers from adopting a cognitive strategy in which they recognized that an 84% contrast is unlikely to be a target (e.g., by directing focal attention away from that location), we interleaved two "decoy" conditions for each cue condition in which the target was made to be 84% and paired with the

same low contrast distracters as presented in the two lowest pedestal conditions. These trials were discarded in subsequent analysis. Each observer ran several hundred training trials before running the experiment which consisted of between 2000-4000 trials. All other details of the experiment were as described above.

Psychophysical contrast-discrimination functions

Contrast-discrimination thresholds were computed separately for each pedestal contrast and each cue condition. During each scanning session, one staircase of forty trials was completed for each cue condition and each pedestal contrast, for a total of 80 staircases per observer (5 staircases per pedestal contrast x 8 pedestals x 2 attention cue conditions). Thresholds were computed from the subject's responses by fitting a Weibull function using a maximum likelihood procedure:

$$p\left(\Delta c\right) = \frac{1}{2} + \left(\frac{1}{2} - \delta\right)\left(1 - e^{-\left(\Delta c/\varepsilon\right)^{\beta}}\right)$$,   [S1]

where $p(\Delta c)$ is the probability of being correct for contrast increment Δc, δ is the fraction of errors at the highest Δc, ε is the midpoint of the abscissa as defined by the position of the upper- and lower-asymptote and it determines the contrast threshold, and β controls the steepness of the psychometric function. Thresholds were computed by reading out of Eq. S1 the contrast (Δc) corresponding to 76% correct. This was repeated for each pedestal contrast, and contrast-discrimination functions were constructed by plotting these thresholds as a function of pedestal contrast.

MRI acquisition and preprocessing

MRI data were acquired on a 3 Tesla Allegra head-only scanner (Siemens, Erlangen, Germany) using an NM-011 head coil to transmit and an NMSC-021 four-channel phased array surface coil to receive (NOVA Medical, Wakefield, MA). Padding was used to minimize observers' head movements; in addition, a bite bar was used with observer 1.

A high resolution anatomy of each observer's brain was used to generate flattened representations of the cortex and to align data across scanning sessions. We acquired three 3D T1-weighted anatomical

volumes (MPRAGE TR 1.5 s, TI 900 ms, TE 3 ms, flip angle 1º, voxel size 1x1x1 mm) in a single scanning session for each observer. These three images were coregistered and averaged. We then used the public domain software FreeSurfer (URL:surfer.nmr.mgh.harvard.edu), to segment the gray matter from these averaged anatomical volumes. All subsequent analyses were constrained only to voxels that intersected with the gray matter.

An anatomical volume with the same slice prescription as the functional images was acquired at the beginning of each scanning session (2D T1-weighted image, MPRAGE slice selective inversion recovery, TR 1400 ms, TI 900 ms, TE 3.79 ms, 14 or 27 slices, voxel size 1.5x1.5x2.5 mm with 0.5 mm gap between slices). We used image-based registration to find the best transformation that aligned these "in-plane" anatomical images to the "canonical" (high resolution) anatomical volume.

Functional scans were acquired with T2*-weighted, gradient recalled echo-planar imaging to measure blood oxygen level-dependent (BOLD) changes in image intensity. We performed retinotopic mapping by acquiring 27 slices with a TR of 1.5 s (TE 30 ms, FA 75º, voxel size 3x3x3 mm, grid size 64x64). We performed the contrast-discrimination experiment by acquiring 14 slices with a TR of 0.8 s (TE 30 ms, FA 57º, voxel size 3x3x3 mm, grid size 64x64). Slices were perpendicular to the calcarine sulcus. The first two volumes in each scan were discarded to allow longitudinal magnetization to reach steady-state.

Functional data for each scan were preprocessed using standard procedures for motion compensation, linearly detrended and high-pass filtered with a cutoff frequency of 0.01 Hz to remove low frequency drifts, and converted to percent signal change by dividing the time-course of each voxel by its mean image intensity over the length of one scan. Time-courses from each session of the main experiment were co-registered using linear interpolation and then concatenated.

Stimulus presentation

Visual stimuli were presented on a 10-bit LCD projector (Eiki LC-XG100, Rancho Santa Margarita, CA; resolution 1024x768 pixels). Observers viewed the stimuli through a mirror (attached to the head coil) as projected on a TechPlex 150 rear-projection screen (Stewart Film Corp., Torrance, CA). The screen was

attached at the end of the magnet bore at a viewing distance of 57 cm, yielding a field of view of 29x22°. Stimuli were generated using Matlab (The Mathworks Inc., Natick, MA) and MGL (URL:http:// justingardner.net/mgl). The projector was gamma-corrected with custom Matlab code using a PR650 SpectraColorimeter (Photo Research, Chatsworth, CA).

Retinotopic mapping

Visual areas V1-hV4 were identified, separately for each observer, based on retinotopic mapping of visual field eccentricity and polar angle (Wandell et al., 2007). Visual area boundaries were drawn by hand by examining coherence, amplitude, and phase maps on flattened representations of the cortical surface. We used published conventions for defining visual field boundaries based on phase reversals of the response to rotating wedge stimuli (Wandell et al., 2007). Data were acquired and analyzed for visual areas V1, V2, V3 and hV4. We note that there is some controversy over the definition of area hV4. We used the definition of hV4 proposed by Wandell et al. (Wandell et al., 2007), because the combined results of our retinotopic mapping and of our stimulus localizer agreed with such definition. Our conclusions would not have differed qualitatively had we adopted any of the other proposed definitions.
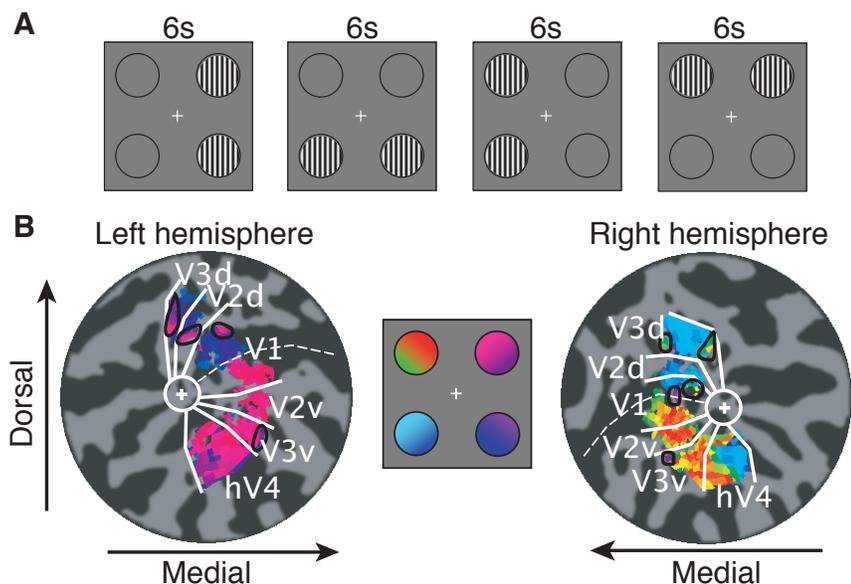
Visual field quadrant localizer

The locations of the visual stimuli in the main experiment (5° in diameter gratings centered at 6° eccentricity along the four 45° diagonals) were mapped using a phased-locked, block-design localizer. Localizer scans (5 minutes and 20 s long, 400 0.8 s TRs) were run at the beginning and end of each scanning session, leading to a total of 30 localizer-scans per observer (20 for observer 2). Each localizer scan started with a grating stimulus presented in the top-right quadrant. Grating stimuli were presented contrast reversing at 5 Hz, and moved in clockwise direction from the top-right to the top-left quadrants, with a period of 24 s and a 50% duty cycle. This stimulus cycle was repeated 13 times in each scan. The first 40 TRs were then discarded to allow for magnetic saturation and steady state response. Data were averaged across all scans and analyzed using the same methods (coherence and phase maps) as for the retinotopic mapping. We restricted each visual area to regions corresponding retinotopically to the stimulus aperture in one quadrant by mapping the phase of the localizer data to the corresponding stimulus

quadrant. All analyses in the main experiment were performed on data sampled from the gray matter portion of these regions using different coherence cutoff thresholds (see online **Supplemental Data** and Fig. S1B and S2B for more details).

Separating BOLD cortical responses to spatially distant stimuli

Our analyses depended on the ability to clearly separate voxels responding to different stimulus locations. In the current study we: (a) tested (in pilot experiments) a range of stimulus parameters (stimulus size, eccentricity, duration and flicker rate) to find those parameters that would allow us to most reliably define independent hemodynamic responses to each stimulus location. (b) Used a phase-mapped localizer to define regions responding to each stimulus location, (c) Repeated all the analyses reported in the main text with different voxel-selection criteria (localizer coherence level, see Experimental Procedures for more details) to ensure that the size of the region used to represent each location would not change our conclusions.

The phase-mapped localizer we used consisted of two 5-Hz counter-phasing flickering gratings that were presented at each of the stimulus locations in sequence (**Fig. S3 A**, see **Visual field quadrant localizer** above for details). Voxels were selected for further analysis if they responded to the stimulus (coherence value 0.7 or 0.5) and intersected the gray matter. Voxels were classified according to which stimulus they responded to (phase of response to

Figure S3. Stimulus localization protocol. A. Trial sequence for one cycle of the phase-map stimulus localizer. **B.** Representative stimulus localizer results in one participant. The left and right panels show a flattened representation of visual cortex in the corresponding hemisphere. White lines define the borders of each visual area as defined by standard retinotopic mapping. The cross indicates foveal convergence. Dark- and light-gray areas represent sulci and gyri respectively. White dashed lines mark the fundus of the calcarine sulcus. Color mapping to stimulus locations is represented in the middle panel. Data are reported at coherence level of 0.5.

localizer, **Fig. S3 B**). Overall, we found very little overlap between phases (different colors, **Fig. S3 B**) that

coded different stimulus locations. Nonetheless, at the coherence levels tested (0.3, 0.5 and 0.7), it was possible to identify some voxels showing a phase inconsistent with the rest of the voxels for the same stimulus location (black ovals). These mislocalized voxels, likely due to partial-voluming artifacts across a sulcus, were removed from further analysis.

fMRI contrast-response functions

The mean fMRI response time-courses were estimated using deconvolution, i.e., linear regression. Specifically, we computed the mean responses for 20.8 s following the stimulus presentations (any response overlaps were assumed to sum linearly), averaged across trials and across voxels in each visual area, but separately for each pedestal contrast and each stimulus-cue combination. The response time-course for the 0% contrast focal-cue condition was used as a baseline and was subtracted from each of the other response time-courses, to remove any non-selective component of the responses, i.e., any component of the responses not dependent on the pedestal contrast or stimulus-cue combination, but time locked to the beginning of each trial. More specifically, we computed a baseline response time-course that consisted of the response to the 0% contrast focal-cue condition at every trial, by convolving the response to the 0% contrast focal-cue condition estimated by deconvolution with a time-course that had a one at the beginning of each trial and zeros elsewhere. We then subtracted this baseline response time-course from the original time-course and recomputed the deconvolution to obtain the baseline subtracted responses. For some of the analyses we also subtracted the response to the 0% contrast in the distributed-cue condition (see Fig. S1A). This resulted in a set of response time-courses that showed a monotonic increase in response amplitudes for each cue condition (Fig. 4). The error bars on these curves corresponded to ±1 standard error of the mean across stimulus repetitions, and were computed by multiplying the inverse of the covariance of the regression matrix with the sum of squares of the residual of the model fit divided by $t–m$, where t is the number of time points in the scan and m is the number of volumes (26) for which the mean response time courses were computed multiplied by the number of different response types (32).

Response amplitudes were extracted from these response time-courses. Specifically, we computed the mean response time-course, averaging across pedestal contrasts and stimulus-cue combinations, and used that as an estimate of the hemodynamic response function for each visual area and each observer.

The response amplitudes were then computed with linear regression, scaling the average response time-course to fit each individual response time-course (corresponding to a single pedestal contrast and stimulus-cue combination). Contrast-response functions were obtained by plotting these response amplitudes (Fig. 4B,C).

The contrast-response functions were fit (nonlinear least-squares) using Eq. 3 (see Fig. 4). There were a total of 32 data points for each visual cortical area: 8 pedestal contrasts X 4 stimulus-cue combinations. These data were fit with 8 free parameters: b_{ft} (baseline response for focal-cue target), b_{fn} (baseline response for focal-cue non-target), b_{dt} (baseline response for distributed cue target), b_{dn} (baseline response for distributed cue non-target), g_r, g_c, s, and q (response-gain, contrast-gain, and exponents for all stimulus-cue combinations). The responses were averaged across trials in which slightly different contrasts were presented so we used the average contrast (averaged across the two stimulus intervals and averaged across trials) when plotting and fitting the contrast-response functions.

Minimizing the potential effects of anticipatory BOLD components

Sirotin and Das (Sirotin and Das, 2009) reported a spatially global hemodynamic activity in V1 of monkeys trained in a periodic fixation task. This activity was present even in the absence of visual stimulation, preceded the beginning of each trial, and did not correlate with either single or multi-unit activity or local field potentials. There has been some debate regarding the nature of this anticipatory hemodynamic activity (Handwerker and Bandettini, 2011; Kleinschmidt and Muller, 2010; Logothetis, 2010). To avoid any potential confound that this might have created for our data analysis, we took the following steps: (1) The effect reported by Sirotin and Das (2009) was shown only for extremely periodic trials. We randomized trial onset so that the observer could not anticipate the timing of the start of the next trial. (2) The hemodynamic activity observed by Sirotin and Das was reported to be spatially global. We used the response to the 0% contrast focal-cue condition as a baseline (i.e. a measurement of any potential spatially-global anticipatory hemodynamic activity) and subtracted that response out from all of the other locations/conditions (see **fMRI contrast-response functions**, above for details). (3) The effect reported by Sirotin and Das was anticipatory, i.e. it started before the beginning of a trial. We randomized the order of the focal-cue and

distributed-cue trials so that if this anticipatory hemodynamic activity emerged it would have been constant for both types of trials. Nonetheless, to test whether our results were robust to any anticipatory activity that was different between the two trial types, we subtracted the response to the 0% contrast distributed non-target stimuli from all responses from the distributed cue trials and subtracted the response to the 0% contrast focal non-target stimuli from all responses from the focal cue trials; hereafter we will refer to this subtraction as "by-trial."   As expected, the by-trial subtraction nearly eliminated the difference between distributed non-target and focal non-target conditions (by definition these two curves started from the same baseline).   The overall shape of the various contrast-response functions was very similar between the original and the by-trial subtractions (**Fig. S1 A.1 and A.2**). The results of the fits of the sensitivity (**Fig. S2 A.1**) and the selection model (**Fig. S2 A.2**) were qualitatively similar with both subtraction procedures, thus demonstrating that removing anticipatory hemodynamic activity separately for the focal-cue and distributed-cue trials did not affect our main conclusions.

Alternate functional forms used to fit contrast-response

To ensure that our conclusions about the sensitivity and selection models were not biased by this particular functional form of the contrast-discrimination functions, we also used two less theoretically motivated equations to fit the data: a skewed gaussian and a polynomial of third order.   The skewed Gaussian,

$$\Delta c(c) = \Delta c_{max} e^{\left(\frac{-\log(c/c_{mean})}{c_{sd} + \omega \log(c/c_{mean})}\right)^2} - e^{-\frac{1}{\omega^2}}, \qquad \text{[S2]}$$

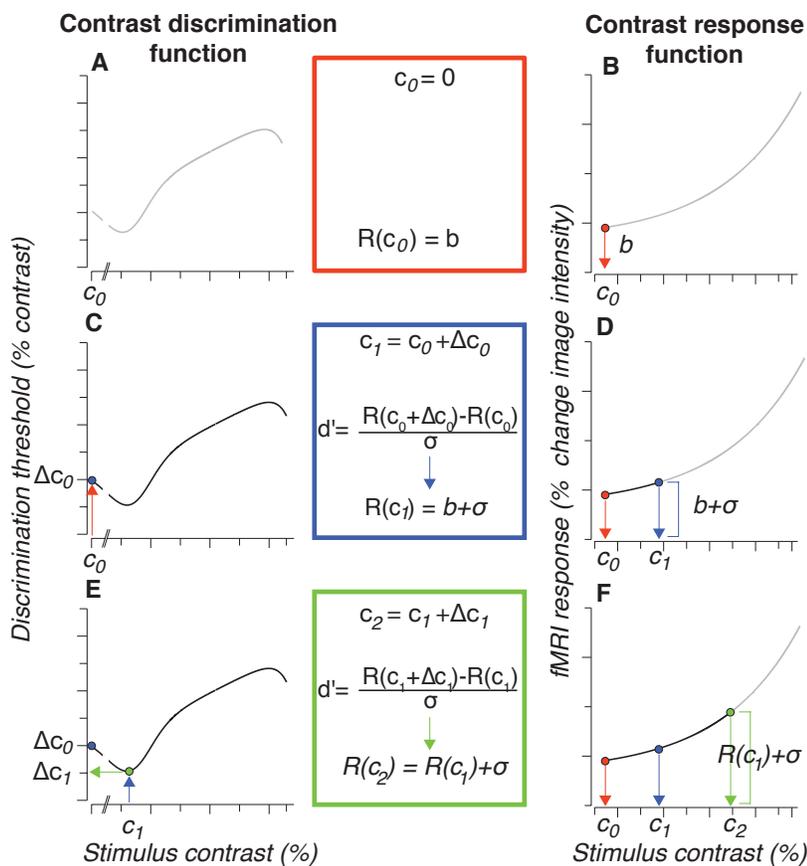had 4 free parameters for each cue condition, where c is the pedestal contrast, Δc_{max} is the maximum measured discrimination threshold, c_{mean} and c_{sd} are the center and the spread of the contrast-discrimination function, and ω controls the skew of the function. The third-order polynomial had 2 free parameters for each cue condition. Both functions fit the data reasonably well and provided a smooth representation of the data.

Fit procedure for the sensitivity model

The sensitivity model predicted contrast-response functions from the measured contrast-discrimination functions by applying the d' equation (**Fig. S4**, blue box). The main parameters of the model were the standard deviation of the noise (σ in the d' equation) and the baseline response (b), which were adjusted to achieve the best fit of the contrast-response functions. To interpolate intermediate values, contrast-discrimination functions were fit with a parameterized form (Eqs. 3 and 4, the exact form of which was not critical for our conclusions – see **Fig. S2C**).

More specifically, For each cortical area we fit the focal-cue and distributed-cue target contrast-response functions (16 data points) using 4 free parameters in the sensitivity model: b_{ft} (baseline response for focal-cue target), b_{dt} (baseline response for distributed cue target), σ_f (noise standard deviation for focal cue), σ_d (noise standard deviation for distributed cue). The contrast-discrimination functions (16 data points) which these fits were based on were interpolated with smooth functions for extrapolating intermediate values needed by the sensitivity model using Eqs. 3-4 (a total of 6 parameters, g_r, s, q, g_c, γ, and ρ, see above, for each contrast-discrimination function).

The fit routine proceeded as follows: for any particular value of

Figure S4. Sensitivity model fit procedure. Contrast-Response Functions (right column, **B**, **D**, **F**) were predicted from fits of contrast-discrimination functions (left column **A**, **C** and **E**) using an iterative procedure that adjusted the noise standard deviation (σ, middle panels) and baseline response, b, to achieve the best fit to the measured contrast-response functions (see text for details) by minimizing the least-squares error between the predicted contrast-response function and the measured data.
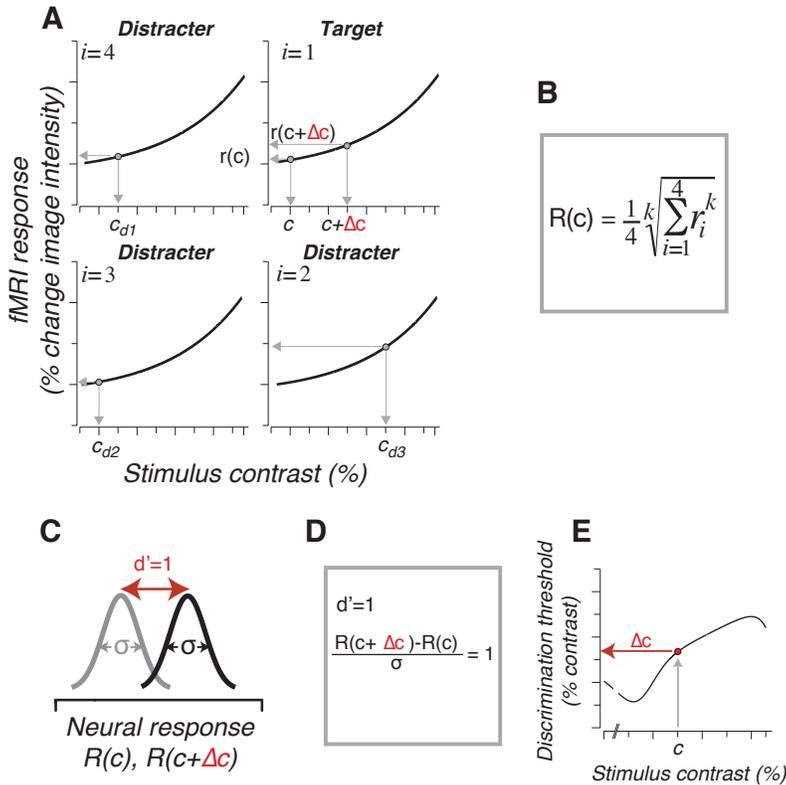
σ and b, the first point on the contrast-response function (c_0=0% contrast) was set to be the baseline response (*b*; **Fig. S4 B**, in red). The next contrast for which a response was estimated was taken from the contrast-discrimination function (**Fig. S4 C**, blue); $c_1 = c_0 + \Delta c_0$ where Δc_0 was the discrimination threshold for contrast c_0. The contrast response at c_1 was then estimated using the equation for d' (blue box) as $R(c_1) = b + \sigma$ (**Fig. S4 D**, blue). The next contrast was then chosen in the same way, by adding to c_1 the contrast threshold value Δc_1 obtained from the contrast-discrimination function (**Fig. S4 E**, green); $c_2 = c_1 + \Delta c_1$. The response at contrast c_2 was then estimated by using the d' equation $R(c_2) = R(c_1) + \sigma$. This procedure was iterated until a full contrast-response function was predicted. σ and b were then adjusted to minimize the least-squares error between the predicted contrast-response function and the measured data.

Fit procedure for the selection model

The selection model predicted contrast discrimination performance based on measured contrast-response functions by first combining responses from different locations together using a pooling equation (Eq. 1) and then applying the equation for d' to these pooled responses. The main model parameters were the standard deviation of the response at each stimulus location (σ) and the degree to which the pooling rule implemented an equally-weighted average to a winner-take-all weighting (k). σ and k were adjusted to provide the best fit of the contrast-discrimination functions. To interpolate intermediate values, a parameterized form of the measured contrast-response function was used (Eq. 3 without the exponent s).

More precisely, for each cortical area we fit the contrast-discrimination functions (16 data points, 8 pedestal contrasts X 2 cues) using the two parameters of the selection model; *σ* (a single noise standard deviation for both the focal cue and distributed cue) and *k* (the exponent for the max-pooling rule which controls whether the pooling rule acts like averaging or max-pooling). The three contrast-response functions (focal-cue target, focal-cue non-target and distributed) were interpolated using a simplified version of Eq. 3 (a Naka-Rushton type equation), which lacked the exponent s, and therefore had 4 parameters *b*, g_r, *q*, and g_c. For the main analysis, this Naka-Rushton equation was fit simultaneously to the contrast-response function data with the rest of the

model. The particular form of the parameterization used to interpolate the contrast-response functions was not essential for our results, nor was it important to simultaneously fit the contrast-response function with the rest of the model. Using the parameterization fit by the sensitivity model (a combination of Eqs. 2-4) or simplified forms based on a skewed-gaussian (4 parameters) and a 2nd order polynomial whose parameters were fixed before fitting the selection model, did not qualitatively change the results (see **Supplemental Data** Fig. S2C).

Figure S5. Selection model fit procedure. A. Idealized representation of the contrast-response functions measured at each stimulus location. Vertical gray arrows show the contrast at each stimulus location for the simulated trial. Horizontal arrows show the mean response. i is an index used in equation **B.** The pooling rule implements a way in which responses can be combined; responses to each stimulus in each interval are summed and elevated to an exponent, *k*. When k is large the largest response dominates the pooling operation, when *k* is 1 responses are averaged together with equal weighting. **c** Pooled response distributions for the two intervals of the trial taken from the contrast-response functions in (**A**) using the pooling rule in (**B**). Contrast-Discrimination Functions (**E**) can be fit by finding the σ and k such that the distributions in **c** have a d' of 1 (the d' value at which thresholds were collected during the experiment).

The fit procedure simulated the performance of an ideal observer on 10,000 randomly created trials, ensuring that the simulated percent correct did not vary more than approximately 1/100th of a percent. The response at each stimulus location and in each interval was taken as a random draw from a gaussian distribution whose mean was set to that of the response for each contrast from the relevant measured contrast-response function and whose standard deviation was the model parameter σ (**Fig. S5 A,C and D**). The four responses in each interval were then pooled using Eq. 1 (**Fig. S5 B**) into a single value. If the larger of these two pooled responses was in the same interval as the increment in contrast, the trial was marked as correct. We then found the Δc that produced correct responses in 76% of the simulated trials

(d'=1, assuming unbiased responses and independent identically distributed gaussian noise) for each pedestal contrast c (**Fig. S5 E**). Finally, σ and k were adjusted to produce the best fit of the contrast discrimination data using least-squares minimization.

Statistical tests

Statistical tests were performed to assess three separate effects of stimulus-cue combination on the contrast-response functions; specifically, whether the response differences could be explained only by (1) a change in g_r (a response-gain change (Williford and Maunsell, 2006)), (2) a change in g_c (a contrast-gain change (Martinez-Trujillo and Treue, 2002; Reynolds et al., 2000; Williford and Maunsell, 2006)), (3) a change in b (a baseline shift (Buracas and Boynton, 2007; Murray, 2008)). Using a nested hypothesis test, each effect (1, 2 or 3) was compared with the "full" model in which all three parameters were allowed to freely vary across stimulus-cue combinations to fit the data. An *F-test* statistic was used to compare the r^2 obtained with the full model to that obtained with the reduced models (1, 2 or 3):

$$F\left(df_1, df_2\right) = \frac{\left(r_a^2 - r_i^2\right)}{df_1} \bigg/ \frac{\left(1 - r_a^2\right)}{df_2} \,,$$
[S3]

where $df_1 = k_a - k_i$, $df_2 = n - k_a - 1$, k_a is the number of parameters for the full model, k_i is the number of parameters for the reduced models, and n is the number of observations (8 contrasts times 4 stimulus-cue combinations). The parameter k_a was set to 14: 4 stimulus-cue combinations times the number of free parameters (g_c, g_r, b) plus the 2 parameters set to be identical across stimulus-cue combinations (s, q). The parameter k_i was set to 8: 4 stimulus-cue combinations times the 1 free parameter (either g_r, g_c or b) plus the 4 parameters fixed across cue conditions (s, q and the other parameters not being tested). We then used the F distribution to estimate the probability that the full model did not differ significantly from the reduced model.

Statistical tests in individual observers

We used a parametric resampling method to test whether the results of the sensory noise reduction and efficient selection fits were statistically significant in each individual observer and visual area. To do so, for each stimulus contrast and cue condition we resampled (with replacement) 100 times from a gaussian

distribution with the mean response amplitude and standard deviation computed (using linear regression) for each of the 4 trial types and 8 contrasts. From each of these simulated data sets, we fit contrast-response functions and estimated the models' parameters, using the same analyses as for the real data. We then compiled the bootstrap distributions of the baseline differences, b_d-b_f, and sensory noise ratios, σ_d/σ_f, and computed p-values based on the number of baseline differences less than 0 and the number of noise ratios less than 1

Model comparisons

Model fits were compared in two different ways, accounting for differences in number of model parameters. One, we computed cross-validated r^2 in which we fit each model on half the data and evaluated the explained variance on the other half of the data (any over-fitting with more model parameters would tend to reduce the r^2 of the model on the left-out data). Second, we computed AIC (An Information Criteria) values with the model residuals and number of parameters using a standard formula which assumes normally distributed errors with constant variance (Burnham, 2002). Smaller AIC values indicate better model fits. AIC is only valid as a relative measure, so we report differences in AIC.

Eye position monitoring

Eye position was monitored using an MRI-compatible, infrared-video eye tracking system (ASL Model-504, Applied Science Laboratories, Bedford, MA). At the beginning of each functional scan, eye position was calibrated; observers fixated a yellow dot that appeared first at the center of the screen and then moved to 5° eccentricity to the left, right, above, and below screen center. Data from this calibration were used to find the best affine transformation (translation, rotation, linear scaling, and linear shear) of the raw eye data to eye position in degrees of visual angle. Stable corneal- and pupil-reflection data were collected during 24 scans. Trial-triggered average (vertical and horizontal) eye position was computed for each trial from stimulus onset to offset, collapsing trials across scans in each scanning session. The standard deviation of the horizontal and vertical eye positions across trials was less then 0.75º of visual angle. The eye position was at central fixation for both focal- and distributed-cue trials. The average eye position in either trial type did not significantly differ from the central fixation (two-tailed 95% confidence

intervals computed by Hotelling T^2 statistics), and there was no systematic difference in average eye position between focal- and distributed-cue trials (single-tailed 95% confidence intervals computed by Hotelling T^2 statistics).

Supplemental References

Buracas, G.T., and Boynton, G.M. (2007). The effect of spatial attention on contrast response functions in human visual cortex. J Neurosci *27*, 93-97.

Burnham, K.P.A., D. R. (2002). Model selection and multimodel inference. A practical information theoretic approach (New York, NY, USA: Springer).

Handwerker, D.A., and Bandettini, P.A. (2011). Hemodynamic signals not predicted? Not so: a comment on Sirotin and Das (2009). Neuroimage *55*, 1409-1412.

Kleinschmidt, A., and Muller, N.G. (2010). The blind, the lame, and the poor signals of brain function--a comment on Sirotin and Das (2009). Neuroimage *50*, 622-625.

Logothetis, N.K. (2010). Neurovascular Uncoupling: Much Ado about Nothing. Front Neuroenergetics *2*.

Martinez-Trujillo, J., and Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron *35*, 365-370.

Murray, S.O. (2008). The effects of spatial attention in early human visual cortex are stimulus independent. J Vis *8*, 2 1-11.

Reynolds, J.H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. Neuron *26*, 703-714.

Sirotin, Y.B., and Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. Nature *457*, 475-479.

Wandell, B.A., Dumoulin, S.O., and Brewer, A.A. (2007). Visual field maps in human cortex. Neuron *56*, 366-383.

Watson, A.B., and Pelli, D.G. (1983). QUEST: a Bayesian adaptive psychometric method. Percept Psychophys *33*, 113-120.

Williford, T., and Maunsell, J.H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. J Neurophysiol *96*, 40-54.